2 Data

The data we will use is publicly available from the government of Ontario at https://www.ontario.ca/data/schoolinformation-and-student-demographics. It provides info about almost 5,000 schools all over ontario, most importantly their location coordinates and students enrolled. Using the location data along with foursquare, a venue search engine, it is possible to search for tutoring services near each of the schools. We will categorize these according to the number of services nearby. We can also collect information about nearby tutoring services if useful or necessary, however we are not for this study.

3 Methodology

Once the data is acquired we must reduce the columns to the ones we will use in our study. We need the school name, location coordinates and enrolment as well as the percentages of low income families and parents with university education. Entries with null school names, coordinates or enrolment values are ignored. Null values in the remaining columns are replaced with the average of the non-null values. We will use the schools in the city of Mississauga to show the process applied. Using the Mississauga school coordinates we search the Foursquare database for ‘tutor‘, ‘math‘, and ‘learning‘ venues at the location of the school. A search radius of 3km was chosen for this example. The results often contained unrelated venues such as those containing the word ‘Matheson‘, a street name. A filter to only include venues with certain subtrings such as ‘mathematics‘ and then exclude those with substrings we do not want, for example ‘adult‘ learning centres. From the resulting venues we collect the number of venues returned for each school as well as a complete list of unique venues returned along with their coordinates for mapping purposes.